

Assessing Visual Quality of 3D Polygonal Models

Abdullah Bulbul, Tolga Capin, Guillaume Lavoué, Marius Preda

Abstract—This survey paper presents recent advances in evaluating and measuring the perceived visual quality of 3D polygonal models. The paper analyzes the general process of objective quality assessment metrics and subjective user evaluation methods and presents a taxonomy of existing solutions. Simple geometric error computed directly on the 3D models does not necessarily reflect the perceived visual quality; therefore, integrating perceptual issues for 3D quality assessment is of great significance. This paper discusses existing metrics, including perceptually based ones, computed either on 3D data or on 2D projections, and evaluates their performance for their correlation with existing subjective studies.

Index Terms—3D polygonal models, quality assessment metrics, perception, subjective evaluation.

I. INTRODUCTION

Technologies underlying 3D computer graphics have matured to the point that they are widely used in several mass-market applications, including networked 3D games, 3D virtual and immersive worlds, and 3D visualization applications [1]. Furthermore, emerging products, such as 3D TVs and 3D-enabled gaming devices, are opening new avenues of opportunity for an enhanced user experience when interacting with 3D environments. Thus, 3D models are emerging as a newly popular form of media [2], usually in the form of 3D polygonal meshes.

3D mesh models are generally composed of a large set of connected vertices and faces required to be rendered and/or streamed in real time. Using a high number of vertices/faces enables a more detailed representation of a model and possibly increases the visual quality while causing a performance loss because of the increased computations. Therefore, a trade-off often emerges between the visual quality of the graphical models and processing time, which results in a need to judge the quality of 3D graphical content. Several operations in 3D models need quality evaluation. For example, transmission of 3D models in network-based applications requires *3D model compression* and *streaming*, in which a trade-off must be made between the visual quality and the transmission speed. Several applications require accurate *level-of-detail (LOD) simplification* of 3D meshes for fast processing and rendering optimization. *Watermarking* of 3D models requires evaluation of quality due to artifacts produced. *Indexing and retrieval of 3D models* require metrics for judging the quality of 3D models that are indexed. Most of these operations cause certain modifications to the 3D shape (see Figure 1). For example,



Fig. 1. Left: original bunny model; middle: simplified; right: smoothed

compression and watermarking schemes may introduce aliasing or even more complex artifacts; LOD simplification and denoising result in a kind of smoothing of the input mesh and can also produce unwanted sharp features. In order to bring 3D graphics to the masses with a high fidelity, different aspects of the quality of the user experience must be understood.

3D mesh models, as a form of visual media, potentially benefit from well-established 2D image and video assessment methods, such as the Visible Difference Predictor (VDP) [3]. Various metrics have thus been proposed that extend the 2D objective quality assessment techniques to incorporate 3D graphical mechanisms. Several aspects of 3D graphics make them a special case, however. 3D models can be viewed from different viewpoints, thus, depending on the application, view-dependent or view-independent techniques may be needed. In addition, once the models are created, their appearance does not depend only on the geometry but also on the material properties, the texture, and the lighting [4]. Furthermore, certain operations on the input 3D model, such as simplification, reduce the number of vertices; and this makes it necessary to handle changes in the input model.

This survey paper presents the latest advances in the field of visual quality evaluation and assessment of 3D graphical models. Sections II and III, respectively, present existing view-independent and view-dependent metrics for objective quality assessment, while Section IV focuses on subjective quality evaluation and presents a taxonomy of existing solutions. Performance evaluations of existing metrics are presented in Section V. Finally, Section VI contains a general discussion and indicates several quality assessment related tools and repositories.

II. VIEWPOINT-INDEPENDENT QUALITY ASSESSMENT

One category of quality assessment metrics directly works on the 3D object space. The quality of a processed (simplified, smoothed, watermarked, etc.) model is generally measured in terms of how “similar” it is to a given original mesh. These similarity metrics measure the impact of the operations on the model. Another possible approach to evaluate the 3D models is to consider 2D rendered images of them according to certain viewpoints; however, viewpoint-independent error

A. Bulbul and T. Capin are with the Department of Computer Engineering, Bilkent University

G. Lavoué is with the Université de Lyon, CNRS, and the INSA-Lyon, LIRIS UMR 5205

M. Preda is with the ARTEMIS Department, Institut TELECOM / TELECOM Management

metrics would be necessary because they provide a unique quality value for a model even if it has been rendered from various viewpoints. Such metrics can be used for comparing compressed models or selecting a level of detail, for example.

A. Geometric-distance-based metrics

The simplest estimation of how similar two meshes are is provided by the root mean square (RMS) difference:

$$RMS(A, B) = \sqrt{\sum_{i=1}^n \|a_i - b_i\|^2}, \quad (1)$$

where A and B are two meshes with the same connectivity, a_i and b_i are the corresponding vertices of A and B , and $\|\cdot\|$ is the Euclidean distance between two points. The problem is that this metric is limited to comparing meshes with the same number of vertices and connectivity.

One of the most popular and earliest metrics for comparing a pair of models with different connectivities is the Hausdorff distance [5]. This metric calculates the similarity of two point sets by computing one-sided distances. The one-sided distance $D(A, B)$ of surface A to surface B is computed as follows:

$$\begin{aligned} dist(a, B) &= \min_{b \in B} (\|a - b\|) \\ D(A, B) &= \max_{a \in A} (dist(a, B)). \end{aligned} \quad (2)$$

As this distance is non-symmetric, the two-sided Hausdorff distance is computed by taking the maximum of $D(A, B)$ and $D(B, A)$ (Figure 2):

$$H(A, B) = \max(D(A, B), D(B, A)). \quad (3)$$

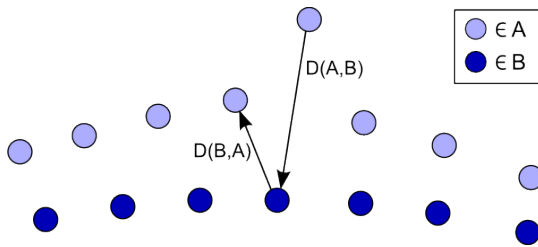


Fig. 2. The Hausdorff distance between two surfaces. The two-sided Hausdorff distance is $H(A, B) = \max(D(A, B), D(B, A))$ [6].

The Hausdorff distance has been used to find the geometric error between a pair of 3D mesh models in the Metro tool by Cignoni et al. [5]. In this approach the mean distance between a pair of meshes is found by dividing the surface integral of the distance between the two meshes by the area of one of the surfaces. The computation of this integral on discrete 3D models requires a sampling method for fast computation. Aspert et al. [7] also propose a sampling implementation of the Hausdorff distance in the MESH tool.

The Hausdorff distance computes the final distance between two surfaces as the maximum of all pointwise distances. Rather than taking the maximum, extensions have been proposed to provide a better indication of the error across the entire surface. Instead of taking the maximum of the pointwise

distances, the average (known as the L_1 norm), the RMS (L_2 norm), and combinations have been proposed [5], [6].

These metrics are well known and widely used; however, even if they can correctly correlate with human judgement in some simple scenarios (see Section V), they usually fail to reflect the perceived quality because they compute a pure geometric distance between a pair of meshes, ignoring the working principles of the human visual system. Hence, several other metrics, using different perceptual principles, have been proposed to better estimate the perceived quality of 3D meshes. These solutions can be categorized as roughness-based, structure-based, saliency-based, and strain-energy-based metrics. Since each of these categories focuses on different aspects of perception, it is unlikely for one of them to estimate the perceived visual quality for all scenarios. In this case, blending metrics of several categories may be a possible solution.

B. Roughness-based metrics

Several solutions evaluate the quality of processed 3D models based on their differences from the original model in their surface roughness (or smoothness). These solutions employ the observation that operations on 3D mesh either introduce a kind of noise related to roughness (e.g., as with quantization or watermarking) or cause smoothing of the surface details (e.g., with level-of-detail simplification for rendering). Roughness is an important perceptual property, as we cannot determine the effect of a small distortion if it is on a rough region of the model, and we can detect defects on smooth surfaces more easily. This perceptual attribute, called the *masking effect*, states that one visual pattern can hide the visibility of another.

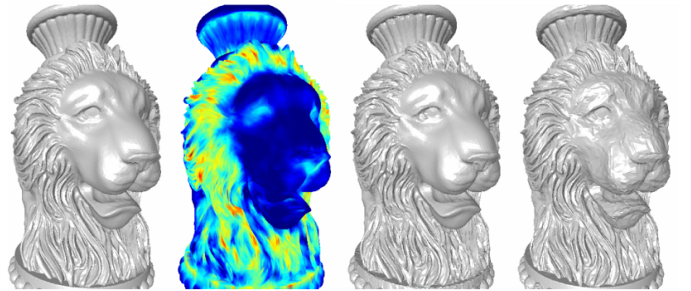


Fig. 3. Roughness map of a 3D model. From left to right: Original model; roughness map: rough regions shown with warmer colors; noise on rough regions; noise on smooth regions. From [8].

Karni and Gotsman propose such a roughness-based error metric to evaluate their mesh compression approach [9]. This metric calculates the Geometric Laplacian of a vertex v_i as follows:

$$GL(v_i) = v_i - \frac{\sum_{j \in n(i)} l_{ij}^{-1} v_j}{\sum_{j \in n(i)} l_{ij}^{-1}}, \quad (4)$$

where $n(i)$ is the set of neighbors of vertex i , and l_{ij} is the geometric distance between vertices i and j . Then the norm

of the Laplacian difference between models M^1 and M^2 is combined with the norm of the geometric distance between the models as follows (v is the vertex set of M):

$$\|M^1 - M^2\| = \frac{1}{2n} (\|v^1 - v^2\| + \|GL(v^1) - GL(v^2)\|). \quad (5)$$

One limitation of this metric is that the compared models must have the same connectivity as the RMS error approach.

Wu et al. [10], for driving their simplification algorithm, examine the dihedral angles of the adjacent faces, considering that a rough surface should have greater dihedral angles. Roughness variation has also been used for quality assessment of watermarked meshes; Gelasca et al. [11] and Corsini et al. [12] measure roughness strength by taking the difference between a mesh and its smoothed version. After computing roughness values for the original and watermarked models, the roughness-based difference is calculated as follows:

$$R(M, M^w) = \log\left(\frac{R(M) - R(M^w)}{R(M)} + k\right) - \log(k), \quad (6)$$

where $R(M)$ is the roughness of the original mesh, $R(M^w)$ is the roughness of the watermarked mesh, and k is a constant to stabilize the numerical results. These roughness-based perceptual metrics [11], [12] have shown to correlate very well with human judgement, particularly in the context of watermarking distortions.

Lavoué proposes a local roughness measure that is able to efficiently differentiate between the different kinds of regions in a mesh: rough parts, smooth regions, and “edge” features, which define border areas between regions [13] (see Figure 3). The proposed measure is based on a curvature analysis of local windows of the mesh and is independent of its connectivity. This measure does not estimate any distance but provides a local roughness estimation that can be used to hide artifacts and could be useful for the design of future quality metrics.

C. Structural distortion-based metrics

Structural distortion-based metrics consider the assumption that the human visual system is good at extracting the structural information of a scene in addition to local properties. Lavoué et al. [14] propose Mesh Structural Distortion Measure (MSDM), based on the work of Wang et al. [15], dedicated to 2D images. Instead of extracting the structural information using luminance in 2D images, this metric uses curvature analysis of the mesh geometry. In this work, a local mesh structural distortion measure (LMSDM) on two local windows x and y of the two meshes is calculated as:

$$LMSDM(x, y) = (\alpha \times L(x, y)^a + \beta \times C(x, y)^a + \gamma \times S(x, y)^a)^{\frac{1}{a}}, \quad (7)$$

with α , β , and γ selected as 0.4, 0.4, and 0.2, respectively, by the authors and with curvature comparison L , contrast comparison C , and structure comparison S computed as:

$$\begin{aligned} L(x, y) &= \frac{\|\mu_x - \mu_y\|}{\max(\mu_x, \mu_y)}, \\ C(x, y) &= \frac{\|\sigma_x - \sigma_y\|}{\max(\sigma_x, \sigma_y)}, \text{ and} \\ S(x, y) &= \frac{\|\sigma_x \sigma_y - \sigma_{xy}\|}{\sigma_x \sigma_y}, \end{aligned} \quad (8)$$

where μ_x , σ_x , and σ_{xy} are respectively the mean, standard deviation, and covariance of the curvature on local windows x and y . Then the MSDM is calculated as follows:

$$MSDM(X, Y) = \left(\frac{1}{n_w} \sum_{i=1}^{n_w} LMSDM(x_i, y_i)^a\right)^{\frac{1}{a}} \in [0, 1], \quad (9)$$

where X and Y are the compared meshes, x_i and y_i are the corresponding local windows of the meshes, and n_w is the number of local windows. a is selected as 3 by the authors, for equations 7 and 9 [14]. This metric has proven to correlate very well with human judgement even in difficult scenarii. The authors propose an improved version of this method in [16].

D. Saliency-based metrics

The metrics described above provide a guarantee of the maximum geometric distance rather than estimating the perceived distance between the models.

In this group of metrics the idea is to give more importance to parts of the meshes that gather more human attention. This type of metric is generally used for mesh simplification such that salient parts of a mesh are preserved in the simplification, as suggested by Howlett et al. [17] and Lee et al. [18]. The salient parts of meshes are determined by utilizing an eye-tracker in Howlett et al.’s work, whereas Lee et al.’s method is more convenient as it computes saliency of a mesh automatically, based on its surface curvature.

Similar to the roughness-based and structural distortion-based metrics, saliency uses the perceptual limitation of the human visual system, and its further use for mesh quality assessment is a research area of great interest.

E. Strain-energy-based metrics

Bian et al. [19] propose a solution based on the strain energy on the mesh as a result of elastic deformation. Mesh models are assumed to be elastic objects; as shells composed of triangular faces of negligible thickness. The assumption is that triangle faces do not bend, and each triangle is deformed along its plane by ignoring any rigid body motion.

The perceptual distance between the two versions of the input model is defined as the weighted average strain energy (ASE) over all triangles of the mesh, normalized by the total area of the triangular faces:

$$SFEM(A, B) = \frac{1}{S} \sum w_i W_i, \quad (10)$$

where w_i are weights for which several strategies are tested in [19] and W_i is the strain energy computed for triangle i .

This model correlates well with human opinion from the subjective experiment conducted by the authors.

F. Attribute-based metrics

Many 3D mesh models contain per-vertex attributes in addition to the vertex position, such as color, normal, and texture coordinates. Also, in sharp creases of the models, there may be multiple normals per-vertex, or there may be several color values on the boundaries, causing discontinuities in the attributes.

As described by Luebke et al. [6], correspondence between vertices on two surfaces is important but is a difficult issue for meshes with different connectivities; it is difficult to compare attribute values from the original surface and a simplified version in a continuous function. Luebke describes an alternative to Hausdorff, called the bijection method. This requires correspondence between vertices in a 2D parametric domain, such as a texture map. This distance is called a parametric distance. Roy et al. [20] propose a metric called Attribute Deviation Metric, that can be used to compare two meshes according to their geometric and appearance attributes (or any other per-vertex attributes). The local deviation of attributes between each point of a mesh and the surface of the reference mesh is calculated using parametric distances.

Pan et al. propose a different approach for quality assessment, calculating the quality of a 3D model according to its wireframe and texture resolutions (Equation 11) [21].

$$Q(g, t) = \frac{1}{\frac{1}{m+(M-m)t} + \left(\frac{1}{m} - \frac{1}{m+(M-m)t}\right)(1-g)^c} \quad (11)$$

Here, m and M are the minimum and maximum bounds of quality, g and t are graphical and texture components scaled into a $[0-1]$ interval, and c is a constant. All coefficients are determined by curve fitting on subjective evaluation data. This metric provides a very good estimation of human judgement as demonstrated in the authors' subjective experiment.

III. VIEWPOINT-DEPENDENT QUALITY ASSESSMENT

Viewpoint-dependent quality assessment metrics estimate the perceptual quality of a 3D model as it is shown on the screen; therefore, these metrics are image-based. Viewpoint-dependent metrics can be classified as: *non-perceptual metrics* and *perceptually based metrics*. The visual system does not matter for non-perceptual approaches; they compute the difference between two images pixel by pixel. Perceptually based metrics rely on the mechanisms of the human visual system and attempt to predict the probability that the human observer will be able to notice differences between images.

A. Non-perceptual metrics

Lindstrom and Turk calculate the RMS image error for mesh simplification [22]. In their work, the meshes are rendered from multiple viewpoints and the quality of the resulting luminance images are measured in terms of their differences from the original image as follows:

$$d_{RMS}(Y^0, Y^1) = \sqrt{\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n (y_{ij}^0 - y_{ij}^1)^2}, \quad (12)$$

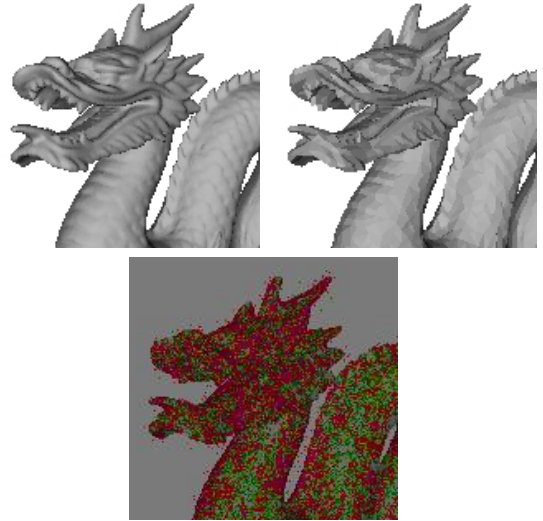


Fig. 4. Left: original image; right: simplified image; bottom: VDP output

where Y^0 and Y^1 are m by n luminance images. The RMS metric is not a good metric for image quality assessment and is seldom used because it is highly affected by a shift or scale, and it does not have a perceptual aspect.

Another quality metric for comparing image quality against a reference image consists in calculating the peak signal-to-noise ratio (PSNR). Using the RMS error shown in Eq. 12, the PSNR for an image with a highest possible intensity value I_{max} can be calculated by:

$$PSNR = 20 \log_{10} \left(\frac{I_{max}}{d_{RMS}} \right) \quad (13)$$

Although PSNR is also widely used for natural images, it is shown to be a poor indicator of image quality [23]. However, according to a report of the Video Quality Experts Group (VQEG), many more-complicated image quality metrics are not significantly better than PSNR [24]. The reasons for this are discussed in a study of Wang et al. [25].

B. Perceptually based metrics

Many 2D metrics incorporate the mechanisms of the human visual system. These metrics generally use the following perceptual concepts: Contrast Sensitivity Function (CSF), which indicates the relation between the visible spatial frequency and different contrast values; and masking, which describes the reduction in the visual sensitivity of a signal upon the existence of another signal.

A popular metric in this category is Daly's Visible Difference Predictor [3]. This metric takes two images as inputs, one of which is evaluated relative to the other; and the output is an image of the perceptual differences between the two images (see Figure 4). The value of each pixel on the output image indicates the detection probability of the difference. The VDP is shown to be a good indicator of perceptually important areas in 3D graphics scenes by the psychophysical experiment of Loughurst and Chalmers [26].

Another well-known metric is the Sarnoff Visual Discrimination Model (VDM) [27] by Lubin. This metric also predicts the detection probability of the differences between a reference image and the evaluated image, as in VDP. The Sarnoff VDM model works on spatial domain whereas VDP works in the frequency domain; VDM works faster but requires more memory. Li et al. [28] compare the two metrics and find that each model has advantageous properties.

Bolin and Meyer modify the Sarnoff VDM model and propose a simpler and faster metric, which incorporates color properties into their 3D global illumination calculations [29]. This metric is preferred for its efficiency. In their subjective experiment with differently simplified 3D models, Watson et al. [30] show that this metric is an effective predictor of fidelity.

Ramasubramanian et al. [31] propose a perceptually based metric that defines a threshold map in which the minimum detectable difference values are stored for each pixel. This metric handles luminance and spatial processing separately, which provides efficiency since it enables pre-computing of the spatial features.

Ramanarayanan et al. [32] introduce a novel concept, the Visual Equivalence Predictor (VEP), which claims that two images are visually equivalent if they give the same impression even though they have visually different parts. This concept makes more sense for computer-generated imagery in which slightly different illumination techniques lead to different images when analyzed pixelwise although the two images have similar fidelity and information. This model takes 3D geometry, material, and illumination properties into account for the equivalency computations. The VEP concept aims to overcome the limitations of the VDP model, which only considers the earliest levels of visual coding, and is therefore too conservative with respect to the kinds of approximations that can be applied in the rendering process.

Visual masking, which describes the reduction in visual sensitivity of a signal upon the existence of another signal, has been used for view-dependent quality assessment of 3D models. Ferwerda et al. [33] investigate the masking effect for computer graphics and extend the VDP model to include color. In their study, a computational model of the masking effect of the used textures on the artifacts of the 3D meshes is developed. This masking effect is predicted on the varying contrast, spatial frequency, and orientation features of the texture pattern and on the polygonal tessellation of the model surface.

Three image-quality metrics based on perceptual color differences are proposed by Albin et al. [34]. These similar metrics find the difference between two images in the LLAB (a modified version of CIELAB) color space. The authors state that these metrics are not complete but only initial attempts at a perceptual quality metric. While the first metric is based on a pixel-by-pixel difference of the images, the second metric gives a single distance value using a Monte-Carlo approach, and the last one is a subdivision-based metric, which gives a rougher difference image compared to the first metric in a shorter time.

IV. SUBJECTIVE EVALUATION OF 3D POLYGONAL MODELS

While automatic metrics are commonly used to predict perceptual quality, relatively few researchers have attempted to measure and predict the visual fidelity of 3D models through subjective experiments. These experiments could be directly used to predict the perceptual quality of 3D models as well as to validate the outcomes of automatic metrics described in the previous sections. Generally, the term “quality” is used to judge how two images (one of them original, the other modified) are “similar” to each other.

A. Experimental measures

Watson et al. study experimental fidelity measures for 3D graphical models [30], and define three of them: *naming time*, which measures the time from the appearance of an object until the observer names it; *rating*, where observers assign a number within a range and meaning determined by the experimenter; and *forced choice preferences*, where observers are shown two or more stimuli, and they choose the stimulus with more of the experimenter-defined quality. The results of this work show that automatic measures of fidelity (e.g., Bolin’s [29], Metro [5], mean squared error (MSE) approaches) are successful at predicting experimental ratings, less successful at predicting preferences, and largely unsuccessful at predicting naming times. On the other hand, when the task is based on comparing different models, ranking is stated to be better than rating the models because the given ratings do not necessarily reflect the perceptual distance between the compared models [35], [36]. The experimental measures used in several user studies can be found in Table I.

B. Experimental design

The parameters used in an experiment are of great importance because they can bias the results significantly, especially for computer-generated stimuli, where almost everything can be controlled. Effective parameters controlled in several quality assessment studies are shown in Table II and listed as follows:

- *Lighting*: The position and type of light source is a crucial element, with a major effect on the viewing conditions. Rogowitz et al. [35] show that models lit from the front result in different subjective scores compared to the same models lit from above. The human visual system has a prior that light is stationary and comes from a left-above orientation [37].
- *Background*: The background may affect the perceived quality by changing the visibility of the boundaries of the model. While a uniform black background is used in several user studies [30] [38], Corsini et al. [12] choose a non-uniform background that fades from blue to white so as not to overestimate the contours.
- *Materials and Shading*: Today, almost all 3D models used in applications have material properties (e.g., texture, normals) and associated complex programmable shaders. On the other hand, most of the subjective evaluations

for verifying perceptual metrics do not take material properties into account; they use only diffuse and smooth-shaded models, mostly to prevent highlight effects [12]. Textures have only been used in the context of substituting geometry with texture [21] [38]. On the other hand, as described above, material properties such as textures introduce the masking effect and hide visual artifacts. Researchers often use models without textures or complex material properties to better control the number of variables influencing the outputs.

- *Animation and Interaction:* To evaluate a 3D model in a fair way, observers should be able to see the models from different viewpoints. This can be achieved by animating the object or viewpoint as in [21], [35], as well as giving free viewpoint control to the user as in [8], [12], [36]. Furthermore, animations affect the perception of the models such that, in the study of Rogowitz and Rushmeier [35], artifacts caused by simplification are less visible when the objects are rotating rather than standing still. The sensitivity of the human visual system is dependent on retinal velocity; the eye's tracking ability is limited to 80 deg/sec [39], which should be taken into account when an experiment includes animation.
- *Type of Objects:* There are several concerns to keep in mind when selecting objects for a subjective experiment. Watson et al. [30] state that evaluation results are different for animal models and man-made artifacts. Further, using abstract objects helps avoid semantic interpretation [38]. Also, the complexity and roughness of the models are important. In a very complex object, simplifications may not be visible and the roughness of a mesh may mask artifacts.
- *Masking:* The object's geometry, roughness, texture, and applied noise or watermarking can mask each other. Lavoue et al. [13] examine the masking effect of noise and roughness; Pan et al. [21] and Rushmeier et al. [38] examine the masking effect of textures on geometry. The masking effect should be considered while designing an experiment.
- *Extent:* The extent, i.e., the display area of the rendered model in pixels, should be large enough to reflect the details of the model. Showing too many items simultaneously may decrease the visibility of the models. The display extents used in several user studies can be found in Table II.
- *Levels:* When an operation (simplification, watermarking, etc.) on meshes is to be tested, the number of the comparison cases and the strengths of the applied operations for each case should be adjusted carefully. Too few levels (compared cases) may not sufficiently reflect the tested operation, whereas a large number of levels may not be feasible, as they would require too many subjects. For simplification case, there are studies using three [30], [35] to seven [36] levels (including the originals) of simplification.
- *Stimuli order:* In comparison-based experiments, stimuli can be shown to the user simultaneously (e.g., side-by-side) or in succession (e.g., first the reference, then

the tested models). When they are shown in succession, enabling users to turn back to the reference model as in the experiment of Rogowitz and Rushmeier [35], allows for a more-detailed comparison. Also, the order and the position of the stimuli should be selected in a way that minimizes the effect of external variables such as observer movements and room's ambient light.

- *Duration:* The duration of which the tested models are shown to the subjects may also affect the results of evaluation.

C. Standards for subjective evaluation

Although no specific recommendation for subjective evaluation of 3D models exists currently, a number of standards, which define the conditions for subjective experiments for other multimedia content (e.g., image and video), could be adapted and used. A well-known standard is the ITU-R BT.500 Recommendation [40], which defines the methodology for the subjective evaluation of image quality. Different experiment methods, such as double-stimulus continuous quality-scale (DSCQS) and simultaneous double stimulus for continuous evaluation (SDSCE) are recommended and grading scales and how to present test materials are outlined. Several of these methods, which may be useful for quality assessment of 3D meshes, are briefly explained below.

- The DSCQS method is recommended for measuring the relative quality of a system against a reference. It has a continuous grade scale which is partitioned into five divisions of equal length, labeled *bad*, *poor*, *fair*, *good*, and *excellent*. Subjects can mark the scale in a continuous manner and then the grades are mapped to a 0-100 interval. The reference and test material are shown twice in succession.
- The SDSCE method is recommended for measuring the fidelity between two impaired video sequences. The stimuli are shown side by side and the grading is continuous.
- The ITU-R BT.500 standard also includes recommendations related to the evaluation of the experiments, such as how to eliminate the outlier data.

A related standard, the ITU-T P.910 recommendation [41], describes subjective assessment methods for evaluating the one-way overall quality for multimedia applications. This recommendation addresses test methods and experiment design, including comparison methods; and evaluation procedures, including viewing conditions and the characteristics of the source sequences, such as duration, kind of content, number of sequences, etc. Subjective evaluation of 3D graphical models as a form of media can benefit from these recommendations.

V. PERFORMANCE EVALUATION

This section is an attempt to provide a performance comparison between the viewpoint-independent and viewpoint-dependent metrics presented in Sections II and III. As an indicator of performance, we restrict to the correlation with human judgment (Mean Opinion Scores coming from the rating of distorted models) captured via the subjective experiments presented in Section IV.

TABLE I
EXPERIMENT METHODOLOGIES OF RECENT SUBJECTIVE EXPERIMENTS ON QUALITY ASSESSMENT

	Masking	Task	Measures	Levels	Purpose of test
Watson01		mesh simp.	rating, preference, naming time	3 (20%, 50%, orig)	
Rogowitz01		mesh simp.	rating	3 (25%, 40%, orig)	to evaluate still images for geo. models
Corsini07		watermarking	rating	E1: 4 mod x 3 wm levels x 3 res. + 4 orig E2: 4 mod x 11 wm types + 4 orig	to fit a metric
Silva08		qual. assesment on mesh simp.	ranking 4 models	10, 20, 27, 35, 43, 50%	to calculate quality of simplified meshes
Pan05	texture-mesh geometry	fitting of subject eval. results to perceptual metric	rating – 5 levels	5 objects, 6 levels of mesh resolution x 3 levels of texture resolution	to fit a metric
Lavoué10	noise-roughness	compression, watermarking, smoothing	rating	E1: orig + 3x noise (0.02, 0.01, 0.005 of bbox) E2: 4 orig + 4x3 noise on smooth + 4x3 noise on rough E3: 4 orig + 4x9 smooth + 4x12 noise versions	to compare different objective metrics
Rushmeier00	texture-mesh geometry	Simplification	rating: 0...100	2 objects x 3 geometry levels (full; 47x reduction in size; 94x reduction in size) x 4 texture levels (none, 512x512, 256x256, 64x64)	to examine possibility of substituting texture for geometry

TABLE II
EXPERIMENT DESIGN OF RECENT SUBJECTIVE EXPERIMENTS ON QUALITY ASSESSMENT

	Lighting	Animation/Interaction	Materials	Background	Object	Extent of Stimulus	simultaneous / successive
Watson01 [30]	oblique	no / no	no	black	man-made vs. animal	591px in width	simultaneous
Rogowitz01 [35]	above, collocated with view	rotating object / no	no		simple vs. complex (vertex count)		successive, can go back
Corsini07 [12]	white point light, top corner of obj. bbox	no / free interaction	diffuse only	blue to white fading		600x600px	
Silva08 [36]		no / free interaction		black	have small num. of vertices and different nature		simultaneous
Pan05 [21]	front	rotation / adjustable rot. speed	textured objects			750px in height	simultaneous
Lavoué10 E1, E2	front	no / free interaction	diffuse only	black	objects with smooth and rough regions		simultaneous
Lavoué10 E3 [8]	white point light, top corner of obj. bb.	no / free interaction	diffuse only	non-uniform	objects from different natures		successive
Rushmeier00 [38]	above, collocated with view	no / no	textured objects	black	abstract objects	370px in height	

The task of comparison is difficult because most of the metrics have been evaluated by a subjective database associated with their own protocols, which sometimes leads to contradictory results. For instance, Rogowitz et al. [35] state that image quality metrics are not adequate for measuring the quality of 3D graphical models because lighting and animation affect the results; on the other hand, based on their experiment with models of varying LODs, Cleju and Saupé [42] claim that image-based metrics are better than metrics working on 3D geometry.

Table III presents the performance evaluation for most of the existing metrics according to the difficulty of the database; indeed if the stimuli come from the same source of distortion (such as a uniform noise addition) even a simple metric will correlate. On the other hand, if the stimuli come from different types of distortions applied on very different 3D models, then it becomes much more difficult to correlate with human

judgement. These data from Table III come from many different sources; for example, in the simplification column, we have synthesized results from Cleju and Saupé [42], Watson et al. [30], and Lavoué [16] using a database from Silva et al. [36]. Most of these results also come from the recent study of Lavoué and Corsini [8], who made a quantitative comparison of several 3D perceptual metrics according to several subjective databases. The rating ranges from ‘-’ for a poor correlation to ‘++’ for an excellent one. ‘?’ indicates that no data is available and ‘N/A’ stands for ‘not applicable’.

Several conclusions can be drawn from this table:

Most metrics are still not applicable for evaluating simplification because they are not able to compare meshes that do not share the same connectivity [9], [14], [19] or the same sampling density [11], [12].

For simple scenarios (a single mesh processing method or a single model), even simple metrics (e.g., Hausdorff) are

TABLE III
PERFORMANCES OF EXISTING METRICS

Metric/Database	Simplification [16], [30], [36], [42]	Unif. noise addition [8], [19]	Watermarking [11], [12], [14], [19]	Masking scenario [14]	Non-unif. noise + smoothing [14]
Hausdorff-based	+	+	+	-	-
RMS	N/A	+	?	-	-
GL [9]	N/A	+	?	-	-
Roughness-based [11], [12]	N/A	++	++	-	+
MSDM [14]	N/A	++	++	+	++
MSDM2 [16]	++	++	?	++	++
Strain Energy [19]	N/A	+	+	-	-
2D Metrics [15], [29]	++	?	?	?	?

able to correlate well with human judgment. However, for more complex scenarios, the roughness-based and structural approaches [11], [12], [14], [16] largely perform better than these simple approaches.

2D metrics seem to perform very well for evaluating simplification but these metrics have never been tested on other scenarios. Actually, a quantitative performance comparison of these image-based metrics against recent 3D perceptually-based ones [11], [12], [14], [16] is still missing. This would be a very interesting task involving any parameters: Which 2D metrics would perform best? How would one choose the 2D views of the 3D models to feed these metrics and how would one combine the scores coming from different views into a single one for the whole 3D model?

VI. DISCUSSION AND CONCLUSION

3D graphical models are emerging as a newly popular form of media, and the importance of quality assessment of these models is only expected to grow. We have surveyed different metrics and approaches for evaluating the visual quality of 3D polygonal models; however, there are several issues to consider in the future.

It is possible to assess the visual quality of 3D models through subjective user tests or automatic measures of quality through view-dependent or view-independent metrics. Regarding the perceived quality directly, subjective evaluation has an advantage over metrics; however, applying user tests is not practical or costly for all applications. In certain applications, such as LOD rendering, quality also has significance for runtime; therefore, there is always a need for better automatic metrics for 3D polygonal models.

Objective metrics are categorized as viewpoint-independent metrics and viewpoint-dependent metrics. Most viewpoint-dependent metrics work on the image space instead of working on the 3D geometry of the models. Although there are important research findings on this issue, we cannot clearly say that one group of metrics is superior to the other.

The nature of the models also affects the perceived quality. An important finding of Watson et al. [30] is that automatic

measures predict the experimental ratings worse for animals (objects with smooth shapes) than man-made artifacts. Also, using familiar objects or abstract objects may change the results due to the semantic interpretations of the objects [38]. Furthermore, usage of textures, which is present in almost all current 3D models, has a great impact on perceived quality since they can mask artifacts on the models [21], [38]. Animation also affects the perceived quality of 3D graphical models [43], [44]. Taking the above factors into consideration, there is a need for more-comprehensive quality metrics that consider these different channels of the 3D model.

Humans' visual system characteristics are considered in many of the metrics in order to better reflect the perceived quality. The adaptation ability of the visual system, the masking effect, and the contrast sensitivity function are the mostly used concepts. In addition to these visibility-related models, attention-oriented metrics, which deal with predicting the highest-attended locations of 3D models, have the potential for further development. Although attention and saliency concepts are studied for 3D models and used in several applications such as mesh simplification [17], [18], [45], there is a need for further work on developing attention-based quality metrics.

This paper has also surveyed various subjective evaluation approaches for 3D model quality assessment. The experiment design has a significant effect on the perceived quality because design decisions, such as the location of light sources, free interaction with the model, or the extent of the model on screen, bias the results. Even if the existing subjective studies seem to have produced relevant results with no error or bias in the protocols, there is a critical need for a real standard for testing 3D graphical models.

Finally, we list some online platforms, tools, and repositories related to 3D model quality assessment. These are generally specialized for a specific application area. For mesh compression, the MPEG 3DG group has initiated an activity on scalable complexity mesh compression to merge the theoretical models and the content used in real applications. The group uses an online platform [46] to be able to deal with 3D graphics objects with various characteristics. For mesh simplification, there are several available tools including QSlim [47] and MeshLab [48]. Another important application area is watermarking as stated by recent studies [11], [12], [14]. Wang et al. [49] have recently presented a benchmarking system for evaluating the 3D mesh watermarking methods. Several repositories and tools, including PolyMeCo [50] and AIM@SHAPE [51], are constructed to ease the validation process of quality assessment metrics by providing tools and common comparison sets. There are also various free, general-purpose tools, providing 3D model comparison (e.g., using geometric distance) such as Metro [5], MESH [7], MeshDev [20]. These platforms and repositories can serve as a tool for future research in this field.

ACKNOWLEDGMENT

The authors would like to thank Rana Nelson for proofreading. This work is supported by the Scientific and Technical Research Council of Turkey (TUBITAK, Project number:

110E029). The bunny and the dragon models are obtained from Stanford University Computer Graphics Laboratory.

REFERENCES

- [1] J. Cooperstock, "Multimodal telepresence systems," *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 77–86, January 2011.
- [2] L. Daly and D. Brutzman, "X3D: Extensible 3D graphics standard [standards in a nutshell]," *IEEE Signal Processing Magazine*, vol. 24, no. 6, pp. 130–135, November 2007.
- [3] S. Daly, "Digital images and human vision," A. B. Watson, Ed. Cambridge, MA, USA: MIT Press, 1993, ch. The visible differences predictor: an algorithm for the assessment of image fidelity, pp. 179–206.
- [4] K. Myszkowski, T. Tawara, H. Akamine, and H.-P. Seidel, "Perception-guided global illumination solution for animation rendering," in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques, SIGGRAPH '01*. New York, NY, USA: ACM, 2001, pp. 221–230.
- [5] P. Cignoni, C. Rocchini, and R. Scopigno, "Metro: Measuring error on simplified surfaces," *Computer Graphics Forum*, vol. 17, no. 2, pp. 167–174, 1998.
- [6] D. Luebke, B. Watson, J. D. Cohen, M. Reddy, and A. Varshney, *Level of Detail for 3D Graphics*. New York, NY, USA: Elsevier Science Inc., 2002.
- [7] N. Aspert, D. Santa-cruz, and T. Ebrahimi, "Mesh: Measuring errors between surfaces using the hausdorff distance," in *Proceedings of 2002 IEEE International Conference on Multimedia and Expo, ICME '02*, vol. 1, 2002, pp. 705–708.
- [8] G. Lavoué and M. Corsini, "A comparison of perceptually-based metrics for objective evaluation of geometry processing," *IEEE Transactions on Multimedia*, vol. 12, no. 7, pp. 636–649, november 2010.
- [9] Z. Karni and C. Gotsman, "Spectral compression of mesh geometry," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques, SIGGRAPH '00*. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 2000, pp. 279–286.
- [10] J.-H. Wu, S.-M. Hu, J.-G. Sun, and C.-L. Tai, "An effective feature-preserving mesh simplification scheme based on face constriction," in *Proceedings of the 9th Pacific Conference on Computer Graphics and Applications, PG '01*. IEEE Computer Society, 2001, pp. 12–21.
- [11] E. Drelie Gelasca, T. Ebrahimi, M. Corsini, and M. Barni, "Objective Evaluation of the Perceptual Quality of 3D Watermarking," in *Proceedings of IEEE International Conference on Image Processing, ICIP, 2005*.
- [12] M. Corsini, E. Drelie Gelasca, T. Ebrahimi, and M. Barni, "Watermarked 3D Mesh Quality Assessment," *IEEE Transactions on Multimedia*, vol. 9, no. 2, pp. 247–256, 2007.
- [13] G. Lavoué, "A local roughness measure for 3d meshes and its application to visual masking," *ACM Transactions on Applied Perception*, vol. 5, pp. 21:1–21:23, February 2009.
- [14] G. Lavoué, E. D. Gelasca, F. Dupont, A. Baskurt, and T. Ebrahimi, "Perceptually driven 3D distance metrics with application to watermarking," in *Proceedings of SPIE Applications of Digital Image Processing XXIX*, vol. 6312, 2006.
- [15] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, april 2004.
- [16] G. Lavoué, "A Multiscale Metric for 3D Mesh Visual Quality Assessment," in *Eurographics Symposium on Geometry Processing*, 2011.
- [17] S. Howlett, J. Hamill, and C. O'Sullivan, "An experimental approach to predicting saliency for simplified polygonal models," in *Proceedings of the 1st Symposium on Applied perception in graphics and visualization, APGV '04*. New York, NY, USA: ACM, 2004, pp. 57–64.
- [18] C. H. Lee, A. Varshney, and D. W. Jacobs, "Mesh saliency," in *ACM SIGGRAPH 2005 Papers, SIGGRAPH '05*. New York, NY, USA: ACM, 2005, pp. 659–666.
- [19] Z. Bian, S.-M. Hu, and R. R. Martin, "Evaluation for small visual difference between conforming meshes on strain field," *Journal of Computer Science and Technology*, vol. 24, pp. 65–75, January 2009.
- [20] M. Roy, S. Fofou, and F. Truchetet, "Mesh comparison using attribute deviation metric," *Journal of Image and Graphics*, vol. 4, pp. 1–14, 2004.
- [21] Y. Pan, I. Cheng, and A. Basu, "Quality metric for approximating subjective evaluation of 3-D objects," *IEEE Transactions on Multimedia*, vol. 7, no. 2, pp. 269–279, april 2005.
- [22] P. Lindstrom and G. Turk, "Image-driven simplification," *ACM Transactions on Graphics*, vol. 19, pp. 204–241, July 2000.
- [23] Z. Wang, H. Sheikh, and A. Bovik, "No-reference perceptual quality assessment of jpeg compressed images," in *Proceedings of IEEE International Conference on Image Processing, ICIP*, vol. 1, 2002, pp. 477–480.
- [24] A. M. Rohaly, J. Libert, P. Corriveau, and A. Webster, "Final report from the video quality experts group on the validation of objective models of video quality assessment," Video Quality Experts Group, VQEG, Tech. Rep., march 2000.
- [25] Z. Wang, Z. Wang, and A. C. Bovik, "Why is image quality assessment so difficult?" in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2002, pp. 3313–3316.
- [26] P. Longhurst and A. Chalmers, "User validation of image quality assessment algorithms," in *Proceedings of the Theory and Practice of Computer Graphics 2004, TPCG'04*. Washington, DC, USA: IEEE Computer Society, 2004, pp. 196–202.
- [27] J. Lubin, *A Visual Discrimination Model for Imaging System Design and Evaluation*, E. Peli, Ed. World Scientific, 1995.
- [28] B. Li, G. W. Meyer, and R. V. Klassen, "A comparison of two image quality models," in *Proceedings of SPIE Human Vision and Electronic Imaging III*, vol. 3299, 1998, pp. 98–109.
- [29] M. R. Bolin and G. W. Meyer, "A perceptually based adaptive sampling algorithm," in *Proceedings of the 25th annual conference on Computer graphics and interactive techniques, SIGGRAPH '98*. New York, NY, USA: ACM, 1998, pp. 299–309.
- [30] B. Watson, A. Friedman, and A. McGaffey, "Measuring and predicting visual fidelity," in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques, SIGGRAPH '01*. New York, NY, USA: ACM, 2001, pp. 213–220.
- [31] M. Ramasubramanian, S. N. Pattanaik, and D. P. Greenberg, "A perceptually based physical error metric for realistic image synthesis," in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques, SIGGRAPH '99*. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 1999, pp. 73–82.
- [32] G. Ramanarayanan, J. Ferwerda, B. Walter, and K. Bala, "Visual equivalence: towards a new standard for image fidelity," in *ACM SIGGRAPH 2007 papers, SIGGRAPH '07*. New York, NY, USA: ACM, 2007.
- [33] J. A. Ferwerda, P. Shirley, S. N. Pattanaik, and D. P. Greenberg, "A model of visual masking for computer graphics," in *Proceedings of the 24th annual conference on Computer graphics and interactive techniques, SIGGRAPH '97*. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 1997, pp. 143–152.
- [34] S. Albin, G. Rougeron, B. Peroche, and A. Tremau, "Quality image metrics for synthetic images based on perceptual color differences," *IEEE Transactions on Image Processing*, vol. 11, no. 9, pp. 961–971, Sep. 2002.
- [35] B. E. Rogowitz and H. E. Rushmeier, "Are image quality metrics adequate to evaluate the quality of geometric objects," in *Human Vision and Electronic Imaging*, 2001, pp. 340–348.
- [36] S. Silva, B. S. Santos, J. Madeira, and C. Ferreira, "Perceived quality assessment of polygonal meshes using observer studies: A new extended protocol," in *Proceedings of SPIE Human Vision and Electronic Imaging XIII*, vol. 6806, 2008.
- [37] I. Howard and B. Rogers, *Seeing in Depth*. Oxford University Press, 2008.
- [38] H. E. Rushmeier, B. E. Rogowitz, and C. Piatko, "Perceptual issues in substituting texture for geometry," in *Proceedings of SPIE Human Vision and Electronic Imaging V*, B. E. Rogowitz and T. N. Pappas, Eds., vol. 3959, no. 1, 2000, pp. 372–383.
- [39] S. Daly, "Engineering observations from spatiovelocity and spatiotemporal visual models," in *IS&T/SPIE Conference on Human Vision and Electronic Imaging III*. SPIE, vol. 3299, 1998, pp. 180–191.
- [40] "ITU recommendation bt.500-10: Methodology for subjective assessment of the quality of television pictures," 2000.
- [41] "ITU recommendation p.910: Subjective video quality assessment methods for multimedia applications," 1996.
- [42] I. Cleju and D. Saupe, "Evaluation of supra-threshold perceptual metrics for 3d models," in *Proceedings of the 3rd symposium on Applied perception in graphics and visualization, APGV '06*. New York, NY, USA: ACM, 2006, pp. 41–44.
- [43] C. O'Sullivan, J. Dingliana, T. Giang, and M. K. Kaiser, "Evaluating the visual fidelity of physically based animations," *ACM Transactions on Graphics*, vol. 22, pp. 527–536, July 2003.
- [44] K. Myszkowski, "Perception-based global illumination, rendering, and animation techniques," in *Proceedings of the 18th Spring Conference on Computer Graphics, SCCG '02*. New York, NY, USA: ACM, 2002, pp. 13–24.

- [45] A. Bulbul, C. Koca, T. Capin, and U. Gdkbay, "Saliency for animated meshes with material properties," in *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization, APGV '10*. New York, NY, USA: ACM, 2010, pp. 81–88.
- [46] MyMultimediaWorld. [Online]. Available: <http://www.mymultimediaworld.com>
- [47] Michael Garland - QSLim Simplification Software. [Online]. Available: <http://mgarland.org/software/qslim.html>
- [48] MeshLab. [Online]. Available: <http://meshlab.sourceforge.net/>
- [49] K. Wang, G. Lavou, F. Denis, A. Baskurt, and X. He, "A benchmark for 3D mesh watermarking," in *Proceedings of the IEEE International Conference on Shape Modeling and Applications*, 2010, pp. 231–235.
- [50] PolyMeCo. [Online]. Available: <http://www.ieeta.pt/polymeco/index.php/home>
- [51] AIM@SHAPE. [Online]. Available: <http://www.aimatshape.net/>